# PERFORMANCE ANALYSIS OF DOUBLE GATE MOSFET USING MONTE CARLO SIMULATION

BY

FAWAD HASSAN ISMAIL

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2009

Urbana, Illinois

Adviser:

Professor Umberto Ravaioli

# ABSTRACT

In this thesis, we explore the performance characteristics, specifically the drain current drive, of the double gate silicon MOSFET device, using MoCa, the Monte Carlo simulator. Drain current performance is analyzed as a result of varying different parameters like oxide thickness, dielectric constant, and misalignment of top and bottom gates. An interesting result is obtained in the misalignment analysis, according to which overlap with source increases the drain current, even in the presence of drain underlap. Misalignment can be tolerable in devices up to a certain extent depending on the application. High-$\kappa$ dielectrics and small oxide thickness are shown to improve the current drive. Comparison is made between quantum-corrected and classical simulation results. Change in potential and concentration profiles in the quantum-corrected simulation is the result of coupling between the Schrödinger and the Poisson equations. The drain current increase compared to a conventional MOSFET of the same dimensions and materials is shown to be significant. Main features of the full band quantum-corrected Monte Carlo simulator are delineated and its significance at the mesoscopic scale is discussed. Finally recent research on electrothermal analysis is reviewed and its importance in relation to the current work is explained. An outline of possible future work is presented for both the simulator and the device.

*To my mother*

# ACKNOWLEDGMENTS

This work would not have been completed without the guidance and encouragement of my advisor, Professor Umberto Ravaioli. His accurate and useful feedback has steered my research and enabled me to learn many things related to my area and the research process. I would also like to thank Mohamed Y. Mohamed for his help and support. His patience in dealing with me, and his availability whenever I needed help, have enabled me to reach my goals in a timely manner. I also owe a debt of gratitude to my colleagues and friends, Mueen Nawaz, Wonsok Lee, and Asfand Waqar for their suggestions on my work. Many thanks to Azeem Sarwar for helping me settle into academia after a long hiatus. Finally, special thanks go to the ECE Publications Office for editing and finalizing the manuscript of this thesis.

# TABLE OF CONTENTS

# CHAPTER 1

# INTRODUCTION

The semiconductor industry is in a state of flux. After experiencing tremendous growth in the past four decades and becoming one of the major industries of the world, it is approaching its saturation point with an uncertain future [1], [2]. Figure 1.1 shows the prediction made by the International Technology Roadmap for Semiconductors (ITRS) 2008 update, according to which physical gate lengths of microprocessor units (MPU) would be reduced to 71% of previous values every 3.8 years instead of 3 years. The ongoing size and speed scaling of the transistor is no longer feasible in its current form and unless some novel way of looking at the problems is discovered, we may witness the demise of Moore's law, the prediction that the number of transistors on a chip would double every 18 months [3], [4].

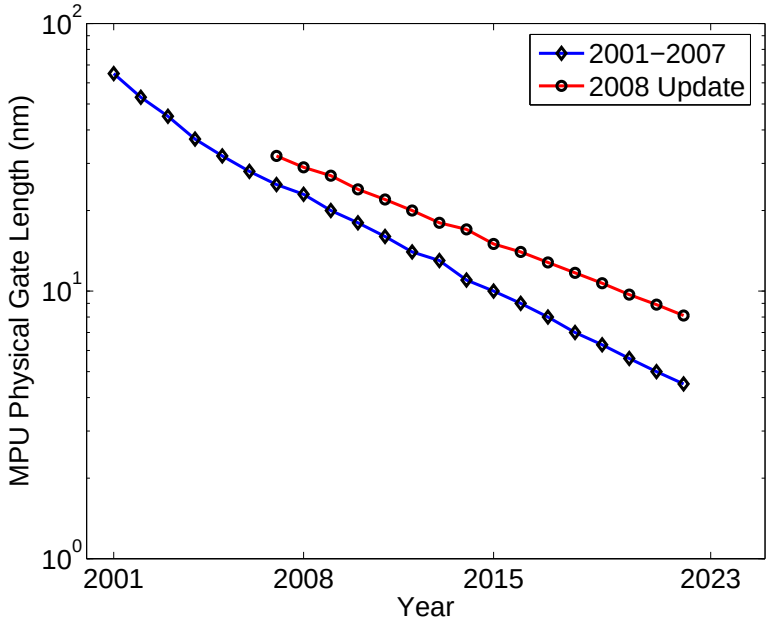


Figure 1.1: Scaling trend reported by ITRS. In 2008, ITRS announced that the scaling trend has slowed down [3], [5].

Researchers continue to devise innovative ideas to prevent this. Computation systems based on DNA, molecular technology, nanotechnology (e.g., nanowires and nanotubes), and quantum mechanics are enjoying vigorous research activity in both academia and industry. However the possibility of further advancement in MOSFET technology cannot yet be precluded [6]. CMOS technology scaling which has worked successfully for the last few decades, sometimes in the face of predictions of failure, is still a promising area of research due to the fact that there is a plethora of solid state phenomena at the atomic scale that need to be understood [7]. New MOSFET architectures like double gate, trigate, and gate-all-around (GAA) have proven to be promising alternatives that can keep the scaling of transistor size and speed from slowing down in coming years [8]. This work will discuss one of these devices, the double gate MOSFET, which is believed to be a superior technology because of its simplicity and compatibility with the existing planar process.

## 1.1 Importance of Device Simulation

Traditionally, semiconductor devices are studied through experiments. But this approach has limitations. We live in the large scale world and can only hypothesize about inner workings of a device by observing its external behavior. Computer simulation, however, is an approach that can enable us to study physical phenomena at the atomic scale. Another advantage of device simulation is the cost factor. Experimenting with novel devices involves fabrication and testing cycles which are expensive in both time and money. With accurate computer simulation, devices can be studied at a fraction of that cost. Device simulation is a necessity in the integrated circuit (IC) fabrication process. The semiconductor consortium has reported upwards of 35% of savings in fabrication costs by using software like TCAD [3].

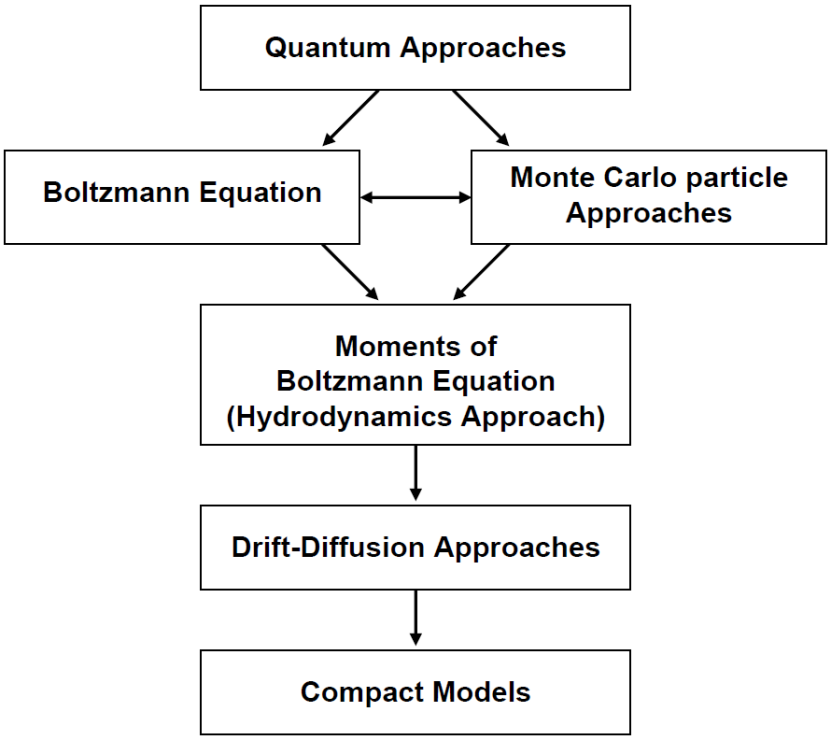


Figure 1.2: Hierarchy of device simulation approaches.

## 1.2 Types of Simulation Approaches

Different approaches have been adopted to simulate semiconductor devices, each with its own benefits and drawbacks. Figure 1.2 shows the hierarchy of semiconductor simulation models in use today. From top to bottom, the models are in decreasing order of computational cost but with decreasing order of accuracy at small scales. The most widely used approach is the drift-diffusion (DD) model which tries to describe the behavior of semiconductor devices through a set of partial differential equations. Simulations based on this model are fast but are accurate only up to a certain scale. Beyond that scale, the hydrodynamics (HD) models can be used that add some parameters to the DD equations to better mimic the reality. But these models are insufficient as well when it comes to simulating high energy, small scale devices. For extremely small devices, quantum modeling is more suitable. However, this model is also computationally expensive, especially if scattering phenomena need to be incorporated. Monte Carlo based schemes, which lie between HD and quantum models, are much more accurate than DD and HD models with

the flexibility of including various atomic scale phenomena of interest like scattering, band-structure etc. The quantum corrected version of this model also compensates for the quantum effects. This model is ideal for devices that are mesoscopic in scale, i.e., devices that start to exhibit quantum phenomena but are not so small that only a handful of atoms are involved. As seen in the figure, running a Monte Carlo simulation is equivalent to statistically solving a Boltzmann transport equation (BTE), which is a seven-dimensional equation that represents the dynamics of charge transport in a semiconductor.

$$\frac{\partial f}{\partial t} + \frac{1}{\hbar}\nabla_k E(\mathbf{k}) \cdot \nabla_r f + \frac{q\mathbf{F}(\mathbf{r})}{\hbar} \cdot \nabla_k f = \left[\frac{\partial f}{\partial t}\right]_{collision} \tag{1.1}$$

where

$f$ : seven-dimensional distribution function (in r, k, and t)

$\mathbf{F}$ : electric-field

$E(\mathbf{k})$ : energy or band-structure

$\mathbf{k}$ : k-space (wave vector)

$\mathbf{r}$ : r-space (real space)

This work focuses on the analysis of a double gate MOSFET device using the technique of full band self-consistent ensemble Monte Carlo with quantum correction. This approach is explained in the next chapter.

# CHAPTER 2

# MOCA—THE MONTE CARLO SIMULATOR

MoCa is an ensemble simulation program [9], [10]. During its running, a large number of charge carriers (a few hundreds to several thousands) are injected into the given device at once and their dynamics are simulated as a result of applied forces. The flow diagram of MoCa is shown in Figure 2.1. Another approach is to simulate only one charge carrier at a time, which is not as efficient.

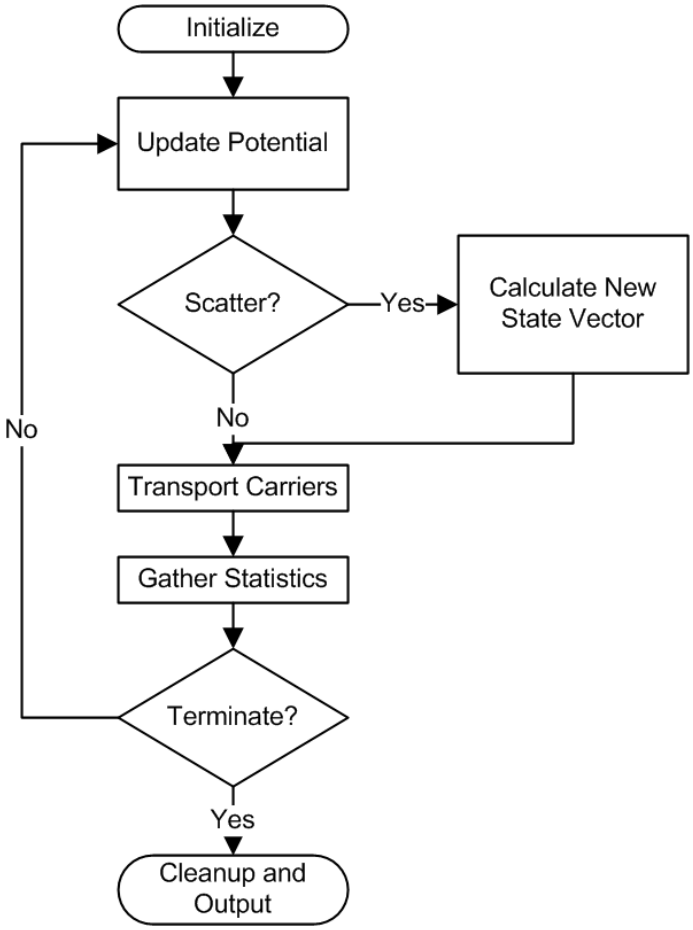


Figure 2.1: Flow diagram of MoCa.

MoCa is capable of performing both bulk-level and geometry-level simulations. To that end, MoCa has been successfully utilized in simulating geometry dependent devices like nanowires and nanotubes. Some of the main features of the simulator are explained below.

## 2.1 Full Band Structure

MoCa takes into account the full electronic band structure of the material involved while performing the simulation. Band structure is the energy versus momentum relationship, also known as the dispersion relation (Figure 2.2). The purpose of the band structure is to calculate the velocities of the particles as they traverse through the device medium under the influence of applied electric field. Simulation programs usually deploy various approximations to the band structure in order to simplify the calculations. Parabolic approximation assumes a quadratic relationship that is valid only for low applied electric fields. For moderate fields, non-parabolic approximation tries to solve the problem by specifying an expression that is non necessarily quadratic. However in order to get accurate results when carriers are subject to high electric fields, full band structure has to employed. MoCa uses tabular data of the full band pre-calculated using the empirical pseudopotential method.

## 2.2 Self-Consistency

The Poisson equation plays an important role in MoCa. It is used to find the classical potential (for a given carrier distribution) in the device. The equation is given as:

$$\nabla \cdot (\epsilon_r \nabla V) = -\frac{\rho}{\epsilon_0} \tag{2.1}$$

where

$V$ : potential

$\epsilon_r$ : relative permittivity of the material

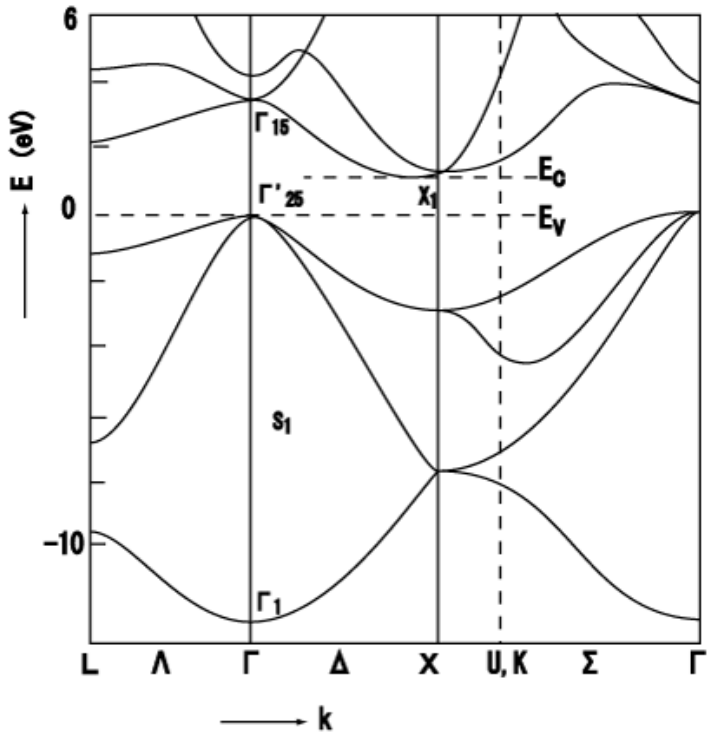


Figure 2.2: Electronic "full" band structure of Si. Inclusion of full band in a device simulation improves the accuracy of the results.

$\epsilon_0$ : permittivity of free space

$\rho$ : charge density

However, for accurate simulation of mesoscopic devices, the nonlinear Poisson equation has to be coupled with the Schrödinger equation in a self-consistent manner because there are some sub-structures, like the inversion layer of a MOS device, that approach the de Broglie wavelength of the electron, and only Schrödinger equation can account for the quantum effects that start to appear at that scale. Following is the time-independent form of Schrödinger equation:

$$-\frac{\hbar^2}{2m}\nabla^2\psi + V\psi = E\psi \tag{2.2}$$

where

$V$ : potential

$E$ : allowed energy level of the particle

$\psi$ : de Broglie wave function of the particle

$m$ : mass of the particle

$\hbar$ : reduced Plank's constant

This coupling becomes more relevant with the ongoing scaling towards the nanometer regime, in order to account for the quantum effects that dominate the behavior at that scale. In MoCa, the coupled system of Schrödinger and Poisson equations is solved. At each step, the Poisson equation is solved and the result is fed into the Schrödinger equation which is solved to put the results back into the next the iteration of the Poisson solver. This process is repeated until convergence is achieved.

## 2.3 Scattering

MoCa uses the scattering scheme that follows from Fermi's Golden Rule and hence is superior to other classical schemes because of its quantum approach. There are myriad scatterings that can be taken into account during the simulation. Most notably the following scattering mechanisms are used: acoustic phonon, optical phonon, impact ionization, impurity, and surface. These scattering mechanisms have pre-determined scattering rates in tabular form which are utilized to have the right proportions of scattering types and events. Here the Monte Carlo technique plays an important role because it is through random sampling that the choice of scattering event is made.

## 2.4 Transport and Time-of-Flight

MoCa statistically solves the BTE that manifests itself in the form of motion of charge carriers through the device until a solution is obtained. The motion of the particles is governed by the following classical equations:

$$\frac{d\mathbf{r}}{dt} = \frac{1}{\hbar}\nabla_k E(\mathbf{k}) \tag{2.3}$$

$$\frac{d\mathbf{k}}{dt} = \frac{q\mathbf{F}(\mathbf{r})}{\hbar} \tag{2.4}$$

The time duration between two scattering events is known as time-of-flight and is calculated using modified constant time technique (MCTT) [11], an improvement upon the constant time technique (CTT) [12]. Since BTE is a classical physics equation, the transport simulation does not take into account the quantum effects.

# CHAPTER 3

# DOUBLE GATE MOSFET

Double gate MOSFET, or DGFET for short, is one of a new generation of devices known as multigate transistors that promise much better scaling capabilities and the potential to stretch Moore's law until roughly the end of the next decade [6]. Other multigate devices being worked on are trigate, gate-all-around, fully depleted silicon-on-insulator (FDSOI), etc. Most of these devices come in both fin-based (called FinFET) and non-fin-based flavors [13]. Double gate has attracted the most attention of all these devices due to the reasons discussed below. DGFET is a simple but clever geometric extension of the conventional MOSFET, in which two gates are used to control the channel instead of one. The 2-D cross-section of the device is shown in Figure 3.1.

The idea of a DGFET was originated in 1983 by Sekigawa and Hayashi when it was called XMOS because of the geometry being similar to the uppercase Greek letter Ξ [14]. It was formally explained by Balestra et al. in 1987 [15]. The first DGFET was fabricated in 1989 [16]. The first Monte Carlo simulation of the device was carried out by Frank et al. in 1992 [17]. Recently the device has gained prominence because semiconductor manufacturers are facing challenges in scaling the conventional MOSFET architecture [18].

## 3.1 Structure

The transistor comes in the following geometrical configurations: planar without substrate, fin-based without substrate, fin-based with substrate. The substrate-less configurations are also called SOI (silicon-on-insulator) transistors. SOI technology is considered to be the precursor to multigate device technology and, if applied to conventional MOSFETs, is an

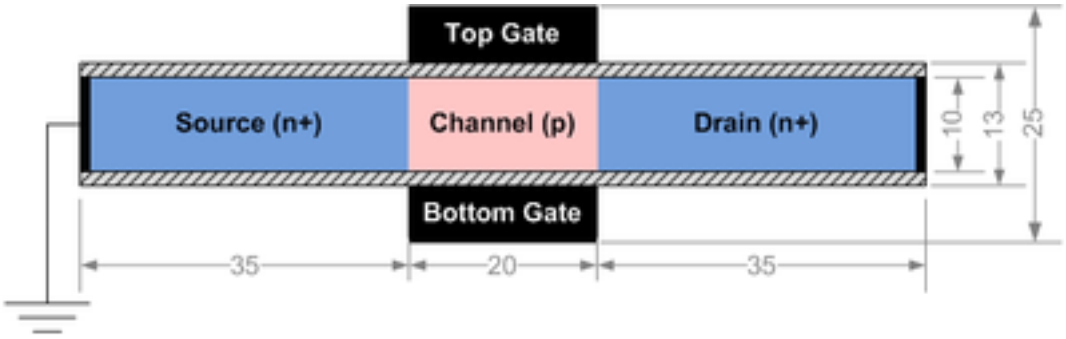


Figure 3.1: Double gate MOSFET geometry to scale. Measurements are in nanometers.

intermediate step toward better scalability. These days SOI is mostly used in conjunction with multigate devices. The channel in a DGFET is kept as thin as possible so as to induce full depletion during the on-state (although partially depleted DGFET is also a possibility). In the fin-based configurations, DGFET geometry is very similar to the trigate geometry with the exception of a much thicker oxide below the top gate so that the top (third) gate is isolated from the device operation.

Planar DGFET is the simplest device in terms of fabrication since it can be constructed using the same processes that are being used by the industry as of today. However, the drawback of this geometry is that the process needs to be extremely accurate in placing the top and bottom gates so as to have no misalignment issues. The fin-based geometry solved the misalignment problem but at the cost of more complicated geometry and need of stringent conditions on the process.

The device can be designed in the connected gate configuration in which only one electrode is provided to the outside world and the potential at that electrode is applied to both gates. The other option is the independent gate configuration in which different voltages can be applied to each gate, making it a four terminal device. This work only explores the two-dimensional cross-section of the DGFET.

## 3.2 The Need for New Architectures

MOSFET scaling has worked for about four decades without much change in the structure, but this is no longer possible because the issues that are

being faced at sub-decananometer scale have new physics and new behavior associated with them. In conventional MOSFETs, the off-current will soon reach a level where more than 50 percent of the power will be consumed while the device is switched off. In other words, static power will surpass the dynamic power. The consequences of this problem are not only that a lot of power is being wasted in the form of heat, but also that it is getting difficult for the logic to operate properly because of diminishing difference between on and off logic levels. The reasons for this leakage problem are multifaceted but can be broadly classified into classical and quantum issues.

Classical issues include DIBL (drain induced barrier lowering), punch-through, and threshold-voltage reduction because of short-channel effects (SCE) [19]. DIBL is the result of capacitive coupling between source and drain due their close proximity. The potential barrier near the source that electrons have to overcome in order to travel towards the drain, is reduced, resulting in increased off current. Punch-through occurs as a result of the coupling between the depletion regions of source and drain in parts of the substrate outside the channel and hence not in control of the gate. This also results in undesirable off current. Finally SCE results when effective depletion charge, i.e., charge in control of the gate bias, is reduced because of source and drain controlled charges coming close to each other. Since the effective depletion charge contributes to maintenance of the threshold-voltage level, the level drops if charge reduces. Another issue related to scaling is the need to scale down supply voltage as well. But it is shown that beyond a certain $V_{dd}$, the static power usage of the device will be greater than the dynamic power, further exacerbating the off state situation [20].

At device dimensions below 50 nm, quantum effects will start playing an increasing role in degrading the device performance. A major quantum phenomenon is tunneling which can be divided into the following kinds. Direct tunneling of carriers into the gate will result because of scaling of the gate dielectric. Direct tunneling from source to drain will be observed as a result of shortening of channel length beyond 5 nm. Field assisted tunneling will be encountered at the gate-drain boundary and will be the result of extremely high electric field in the channel close to the drain.

## 3.3 Salient Features of Double Gage MOSFET

DGFET has advantages that set it apart from conventional MOSFETs and even FDSOI transistors. The absence of the substrate results in elimination of punch-through. It also results in electric field lines terminating at the bottom gate and hence only the field lines that are directed from drain to source have any negative effect on the performance. It is shown that DGFET exhibits reduced threshold-voltage roll-off [15]. Threshold voltage is reduced because of SCE and DIBL according to the following relation:

$$V_{TH} = V_{THo} - \alpha \frac{\epsilon_{Si}}{\epsilon_{ox}} EI(V_{bi} + V_{DS}) \tag{3.1}$$

where $V_{THo}$ is the long-channel threshold-voltage, $V_{TH}$ is the short-channel threshold-voltage, $\alpha$ is a constant, $epsilon_{Si}$ is the dielectric constant of silicon, $\epsilon_{ox}$ is the dielectric constant of the gate oxide, $V_{bi}$ is the built-in voltage between source and drain, and $V_{DS}$ is the applied bias between them. Keeping all other parameters constant, threshold voltage is reduced by increasing the factor EI (electrostatic integrity) which is a function of channel and oxide thicknesses. For a conventional MOSFET the EI relation is given by the VDT (voltage doping transformation) model as:

$$EI = \left[1 + \frac{x_j^2}{L_{el}^2}\right] \frac{t_{ox} t_{dep}}{L_{el}^2} \tag{3.2}$$

EI increases if channel length $L_{el}$ is decreased. In order to limit the increase in EI, the thicknesses of oxide and depletion (in the numerator) have to be decreased. There is a limit to how small these thicknesses can be made. Now the EI relation for a DGFET is given by:

$$EI = \left[1 + \frac{(t_{Si}/2)^2}{L_{el}^2}\right] \frac{(t_{ox}/2)(t_{dep}/2)}{L_{el}^2} \tag{3.3}$$

In other words, all the terms in the numerator are reduced by half, hence relaxing the need to reduce the thicknesses to a very small scale. This also means that for the same channel length, and oxide and channel thicknesses, EI of a DGFET is a fraction of that of a conventional MOSFET.

Furthermore, DGFET depicts a phenomenon called volume inversion that improves the carrier mobility. Volume inversion is the result of close proximity of both gates resulting in a quantum potential well for the charge

carriers in the channel. This potential well restricts the movement of charges to within the volume instead of piling up at the interfaces. This results in reduced interface scattering and hence increased mobility. Volume inversion is shown in Figure 4.5 on page 21.

# CHAPTER 4

# THE SCHRÖDINGER-POISSON COUPLING

In order to transform a classical MoCa to a quantum or nearly quantum MoCa, some kind of quantum correction technique needs to be applied. The major techniques in use are the Wigner formulation, the path-integral formulation, and the Schrödinger formulation—the last one utilizes the Schrödinger equation directly and is being used in MoCa [21], [22]. As a result of the correction, MoCa is able to consider short channel effects (SCE) like size-quantization and tunneling. The Schrödinger formulation based correction can be summarized by the following equation:

$$V_{schr}(z) = -k_B T log(n_q(z)) - V_p(z) + V_0 \tag{4.1}$$

where

$z$ : direction perpendicular to the interface

$V_{schr}$ : quantum-correction in the potential

$V_p$ : potential from the solution of the Poisson equation

$V_0$ : arbitrary reference potential

$n_q$ : charge density obtained from Equation (4.3)

## 4.1 Schrödinger-Poisson Solver

An independent tool [23] is used to isolate the cause of quantum correction in MoCa, i.e., the coupling between the Schrödinger and the Poisson equations [24], [25]. We consider a 2-D version of Equation (2.2) for the DGFET structure (Figure 3.1):

$$\Big[-\frac{\hbar^2}{2m_x^v}\frac{\partial^2}{\partial x^2} - \frac{\hbar^2}{2m_y^v}\frac{\partial^2}{\partial y^2} + V(x,y)\Big]\psi_j^v(x,y) = E_j^v\psi_j^v(x,z) \tag{4.2}$$

At each iteration step of the scheme, we solve the above equation three times, for each equivalent valley pair $v$. Charge density to be used in the next iteration of Equation (2.1) is calculated using the following:

$$\frac{\rho}{q} = n_q = \sum_v \frac{1}{\pi}\sqrt{\frac{2m_z^v k_B T}{\hbar^2}} \sum_n \psi_n^v \psi_n^{v*} \mathcal{F}_{-\frac{1}{2}}\left(\frac{E_F - E_n}{k_B T}\right) \tag{4.3}$$

The value of $\rho$ is then used in Equation (4.6) to get the potential $V$ and this procedure is repeated until convergence is achieved.

### 4.1.1 Setup

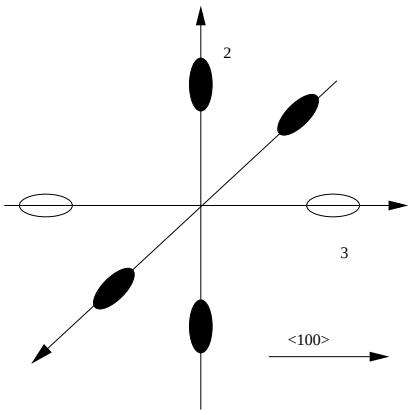


Figure 4.1: Conduction band model for silicon.

The DGFET structure is oriented in the (100) direction using the above setup. We assume a simple 6-valley bandstructure for silicon as shown in Figure 4.1. Each valley corresponds to a constant energy surface, and can be captured by a parabolic effective mass equation. This model allows us to capture the mass anisotropy in silicon. The valley pair pointing in the (100) direction have a mass $m_l = 0.91m_0$ and the transverse mass is $m_t = 0.19m_0$. The two valleys in this direction are degenerate. The other valley pairs also have the same longitudinal and transverse mass along their respective axes.

The rectangular space of the channel with top and bottom oxides is taken into account as a 2-D structure. The top and bottom of that rectangle are

the gate contacts where we apply a fixed potential. The thicknesses of channel and oxide are varied for different configurations of the device.

We use finite difference approximation to discretize the physical space as mentioned in the last paragraph. Assuming the same value for $\Delta x$ and $\Delta y$ (let's call it $a$), we use the five-point stencil formula for the 2-D double derivative as follows:

$$\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \equiv \frac{-u_{i-1,j} - u_{i,j-1} + 4u_{i,j} - u_{i,j+1} - u_{i+1,j}}{a^2} \tag{4.4}$$

### 4.1.2 Numerics

Once the discretization is done, our physical problem is transformed into a numerical problem as follows. The Schrödinger equation (4.2) and the Poisson equation (2.1) are represented by an eigenvalue problem and a linear system problem respectively:

$$\mathbf{A_1 y} + \mathbf{diag(x) y} = \lambda \mathbf{y} \tag{4.5}$$

$$\mathbf{A_2 x} = \mathbf{b} \tag{4.6}$$

where

$\mathbf{A_1}$ : finite-difference form of the second order derivative of the wavefunction

$\mathbf{A_2}$ : finite-difference form of the second order derivative of the potential

$\mathbf{y}$ : wavefunction vector

$\mathbf{x}$ : potential vector

$\lambda$ : energy level scalar

$\mathbf{b}$ : charge density vector

Using Equation (4.4) and the fact that our $\mathbf{x}$ is 2-D in space, the form of $\mathbf{A}$ is as follows:

$$\mathbf{A} = \begin{bmatrix} -4 & 1 & 0 & \cdots & 0 & 1 & 0 & & \cdots & 0 \\ 1 & -4 & 1 & & & & 1 & & & \\ 0 & 1 & -4 & 1 & & & & \ddots & & \\ \vdots & & 1 & -4 & \ddots & & & & & \\ 0 & & & \ddots & \ddots & & & & & \vdots \\ 1 & & & & & & & & & \\ 0 & 1 & & & & & & & & \\ & & \ddots & & & & & & & \\ \vdots & & & & & & & & & \\ 0 & & & & \cdots & & & & & \end{bmatrix} \tag{4.7}$$

As we can see, this is a five-band matrix with three adjacent diagonals in the center (like a tridiagonal) and upper and lower diagonals at a spacing which is the number of grid points in one dimension. So for a 10-by-10 grid, each of the upper and lower diagonals would be 10 diagonals apart from the main diagonal. This means that for an m-by-n grid, the size of the vectors in 4.6 and 4.5 would be m-by-n and the size of the matrices would be m-by-n squared (e.g., for a 10-by-10 grid, $\mathbf{x}$ and $\mathbf{y}$ would have 100 elements and $\mathbf{A_1}$ and $\mathbf{A_2}$ would be 100-by-100 matrices). This apparent $\mathcal{O}(n^4)$ complexity of the problem can be mitigated by the fact that only sparse matrices are involved and also only a small number of eigenvalues from 4.5 are required to proceed with calculations. Hence efficient algorithms can be utilized which exploit these facts and speed up the process.

The application of Dirichlet boundary conditions, as long as they are zero, does not require any change in the matrix or the numerical equation. But for Neumann boundary conditions, the absolute value of the appropriate elements on the diagonal needs to be decremented by one, e.g., –3 instead of –4, or 3 instead of 4. Those appropriate elements depend on where we are applying the Neumann boundary conditions. For example, if they are applied to right and left edges of the 2-D physical space, the elements would be 1st, $n$'th, $(n+1)$'th, $2n$'th, $(2n+1)$'th, and so on.

In order to find the carrier concentration from $\psi_n$ and $E_n$ using Equation (4.3) we need to calculate the Fermi integral of order half, for which an integration algorithm is used.

### 4.1.3 Electron Concentration

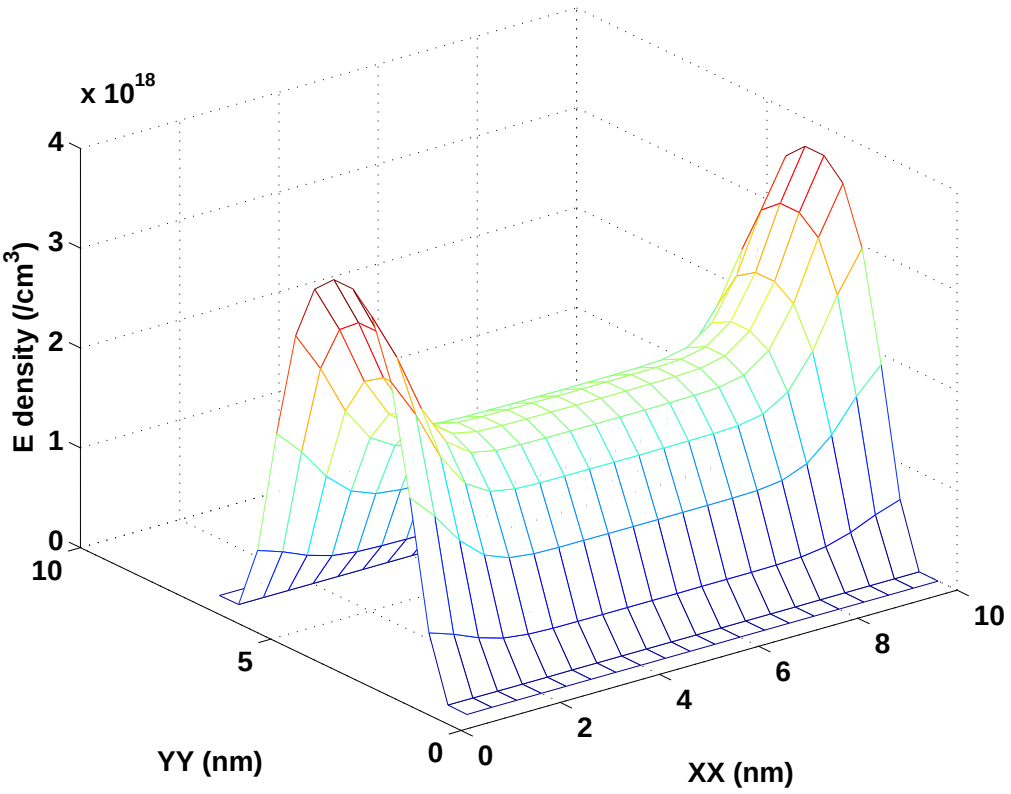


Figure 4.2: Electron density (or concentration) in the 2-D grid used for the solver. YY refers to axis along the channel thickness. XX refers to axis along the channel length.

The electron concentration for DGFET is shown in Figure 4.2. For channel thickness of the order of a few nanometers, electron concentration peaks move away from the edges until both peaks are superimposed on each other in the center of the channel. The reason is that for small dimensions the channel acts as a finite potential well and restricts the probability of electrons at the edges. This results in an increase in effective oxide thickness and consequently a reduction in drain current as will be seen in Figures 5.2 and 5.3 in the next chapter.

The effect of variation of channel thickness is also shown in Figure 4.3 where, for thicker channels, two concentration peaks are separate from each other but merge when thickness is reduced.

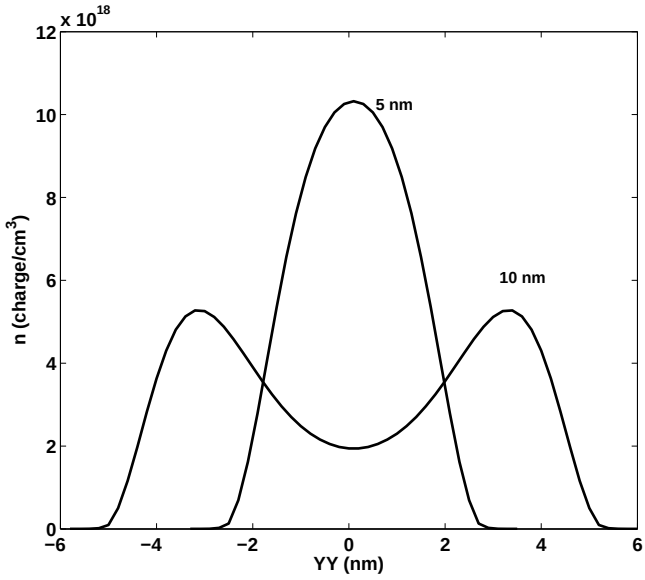


Figure 4.3: Effect of variation in channel thickness on electron density.

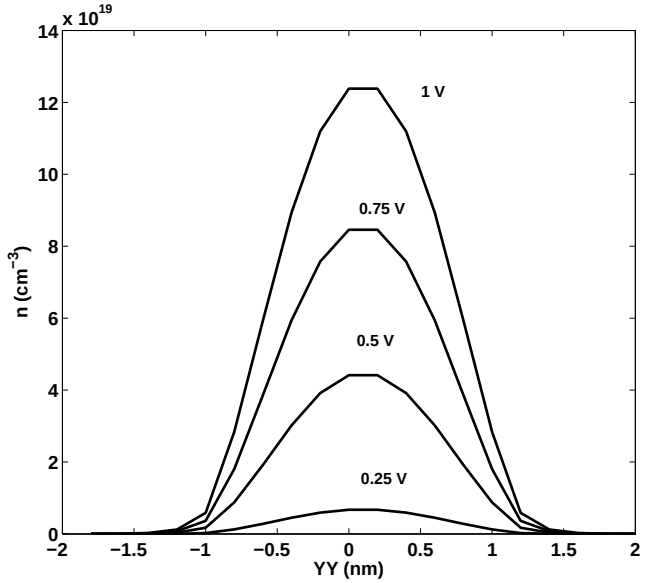


Figure 4.4: Effect of variation in gate voltage on electron density.

Electron concentration versus gate voltage is plotted in Figure 4.4 showing the increase in the availability of charge carriers for higher gate voltages. This results in higher drain currents according to the conventional MOSFET theory.

## 4.2 Comparison Between MoCa Quantum Correction and Schrödinger-Poisson Solver

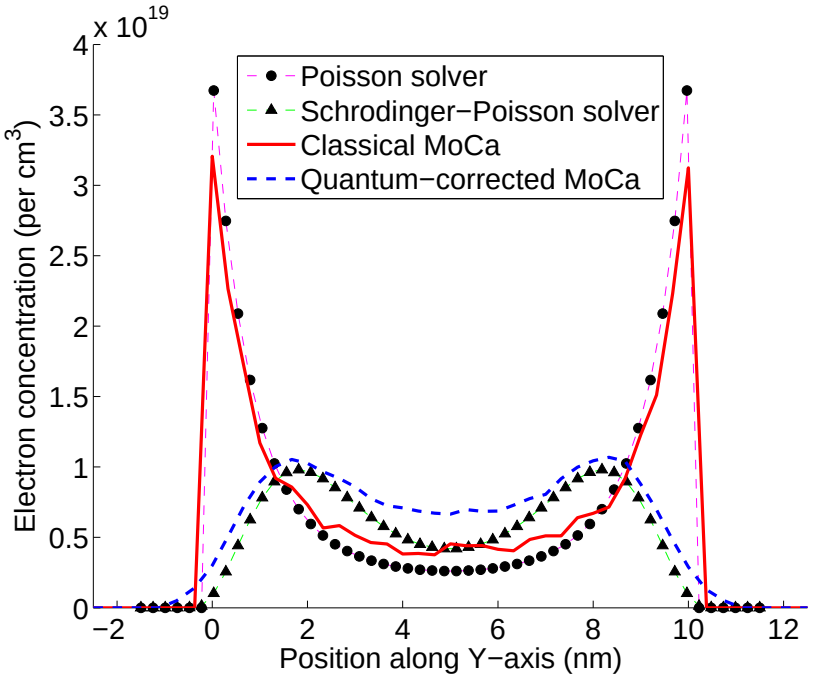


Figure 4.5: Electron concentration in classical and quantum-corrected computations. It is the coupling between Schrödinger and Poisson equations that results in shifting of charge peaks inside the volume, because for small channel thickness, the region acts as a finite potential well.

The electron concentration results of the independent solver are compared with that of the MoCa simulation for the same DGFET structure. In order to reproduce the classical analog of the independent solver, numerics dealing with the Schrödinger equation were removed. The results for electron concentration, calculated by the Poisson-only solver, are plotted in Figure 4.5 along with that of the Schrödinger-Poisson solver. Also plotted in the figure are the electron concentrations as a result of classical as well as quantum corrected MoCa simulations.

The agreement between the two approaches is visible. The classical solution of the Poisson equation gives peak concentration at the interfaces while if the Schrödinger equation is included in each iteration, the peaks move into the volume. The phenomenon is easily observable in MoCa where classical results would correspond to absence of Schrödinger equation considerations. By isolating the solver it is easy to see that the

self-consistent calculation of the two equations is solely responsible for the shift of maximum electron concentration away from the interface. The slight mismatch between the solver and MoCa versions of the plot is mainly because the solver is based on effective mass approximation whereas MoCa includes the full band structure in its calculations.

# CHAPTER 5

# SIMULATION RESULTS

## 5.1 Classical versus Quantum I-V Characteristics

The simulations for the comparison of classical versus quantum I-V characteristics are performed on a Si DGFET with source and drain lengths of 35 nm and gate length of 20 nm. The channel thicknesses used are 10 nm and 5 nm while the oxide thickness is kept at 1.5 nm for both gates. All simulations are done on a 2-D structure; the third dimension of the device is assumed to be 1 $\mu$m. In other words the current values in this work can be taken as mA/$\mu$m for arbitrary length in the third dimension. The channel is moderately doped with $10^{15}$ cm$^{-3}$ while the source and drain are heavily doped with $10^{20}$ cm$^{-3}$. Gate bias was symmetric, i.e., the same voltage on both gates, and was varied between 0.3 and 0.8 V. Drain bias of 0.1 to 1.0 V was used with respect to the grounded source. The simulation was run with 50 000 ensemble particles and time-step of 1 attosecond was used. A total of 0.8 million iterations were carried out for proper convergence of the current value at every combination of drain and gate bias. The convergence rate of two random simulations is shown in Figure 5.1 and we can see that the residual is small enough at the end of the simulations.

The I-V curves for both 10 nm and 5 nm channels is shown in Figures 5.2 and 5.3 respectively. In both cases, quantum corrected currents are lower than the classical counterparts. The reason is the increased influence of quantum phenomena at smaller channel lengths. Also, the 5 nm channel currents are smaller than those of 10 nm channels, which can be attributed to reduced carrier concentration because of the narrowing of the channel and hence reduced mobility.

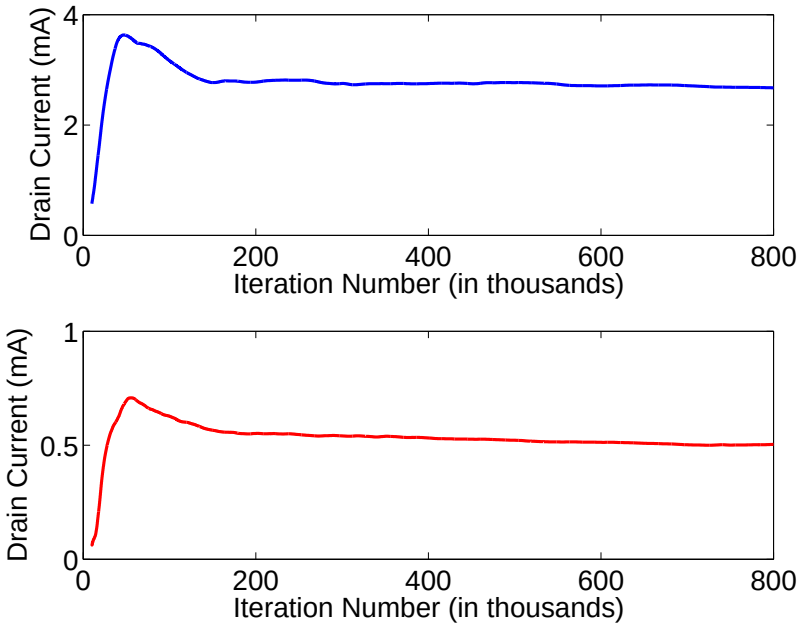


Figure 5.1: Convergence results of two random simulations. 800 000 iterations of MoCa give a fairly converged current value.

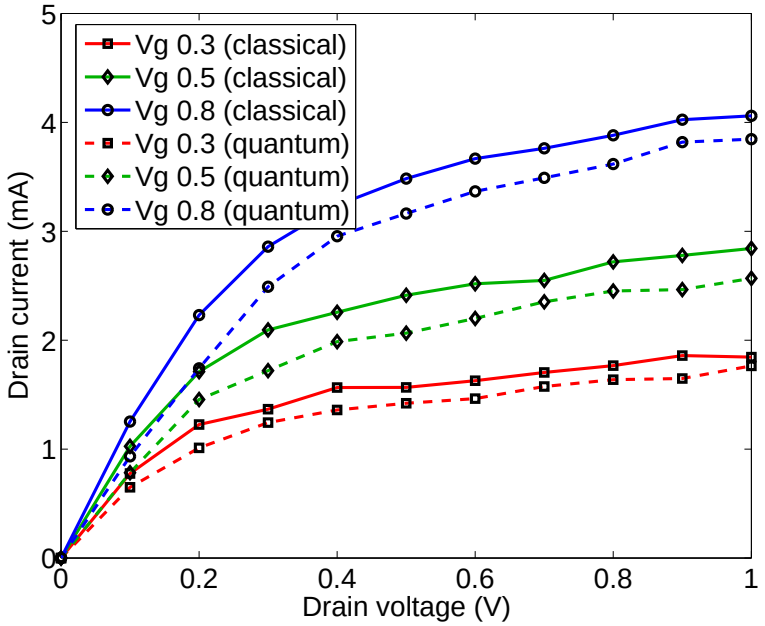


Figure 5.2: I-V characteristics of a DGFET at 20 nm channel length and 10 nm channel thickness. Lower values after quantum-correction are due to accurate estimate of inversion.

## 5.2 Potential and Concentration Profiles

The potential profiles along two cross-sections, horizontal and vertical in reference to Figure 3.1, are plotted in Figures 5.4 and 5.5 respectively. The

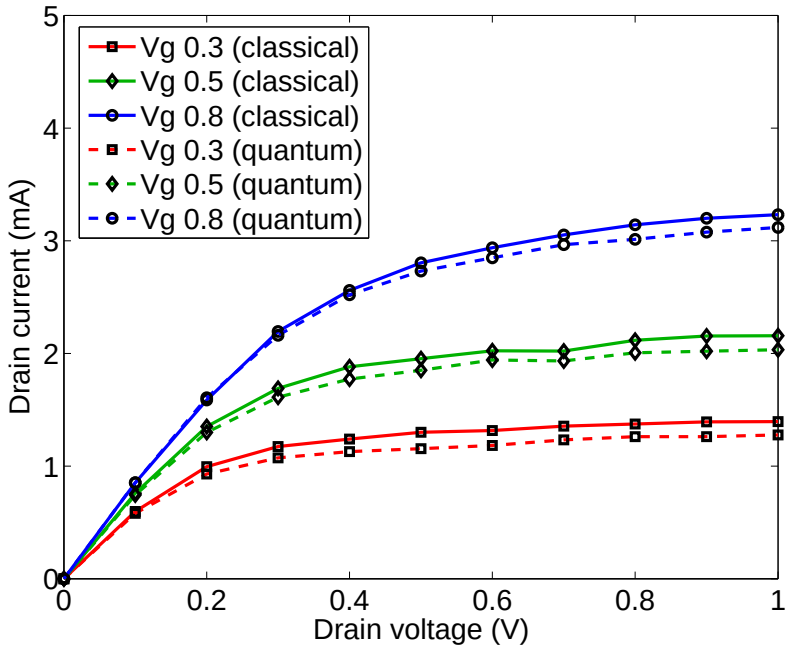


Figure 5.3: I-V characteristics of a DGFET at 20 nm channel length and 5 nm channel thickness. Very small thicknesses result in increased scattering and hence reduced current drive.

horizontal cross-section shows the DIBL effect at 10 nm, reducing the potential barrier at the source end. The vertical cross-section shows the modified potential after the application of quantum-correction. The inversion of the potential curve is the reason behind volume inversion which will be discussed in the next section. The contour of potential throughout the channel is shown in Figure 5.6.

The contour of electron concentration throughout the channel is shown in Figure 5.7. As we can see, concentrations are higher around two imaginary horizontal lines away from the top and bottom edges. The other noticeable point is the higher concentration at the source end compared to the drain end, also visible in Figure 5.8. This can be explained by the fact that the source acts a reservoir of electrons, which enter the channel and travel towards the drain. At the drain end they pass by very quickly because of the very high electric field at the drain end. Secondly, the bias between gates and source is maximum at the source end and hence results in an inversion which is much stronger than that at the source end where the potential difference between gates and drain can be as low as zero. This can be seen in Figure 5.9.

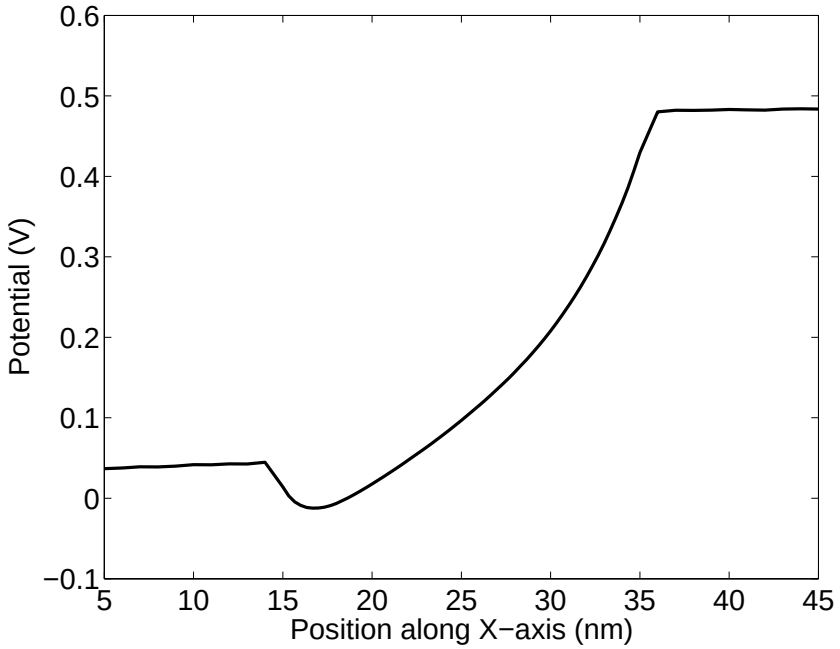


Figure 5.4: Potential profile along the length of the device. Left edge represents source, followed by channel, and then drain. DIBL because of short channel length is visible at 16 nm. For long channels, this value would be lower.

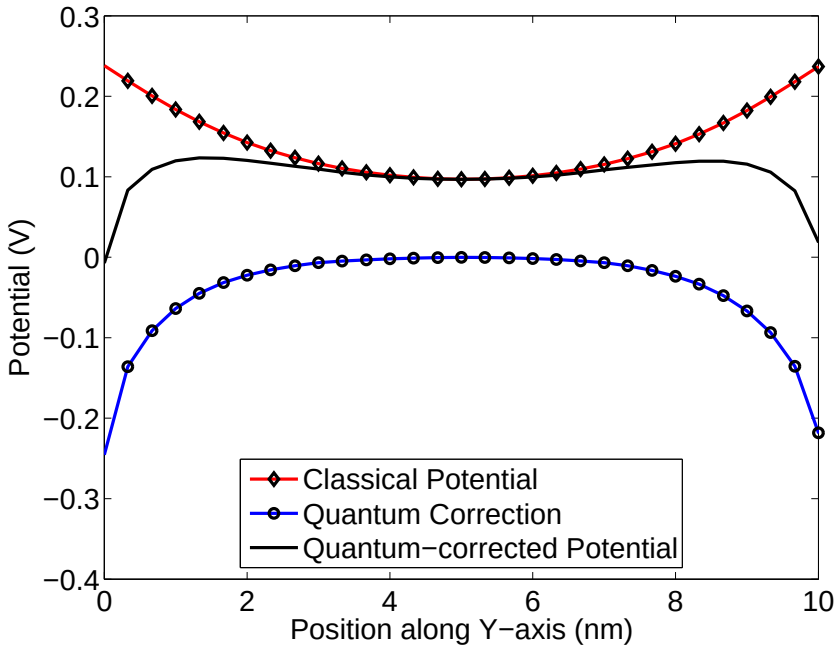


Figure 5.5: Potential across channel thickness. Quantum correction modifies the shape of the potential accurately reflecting the volume inversion phenomenon.

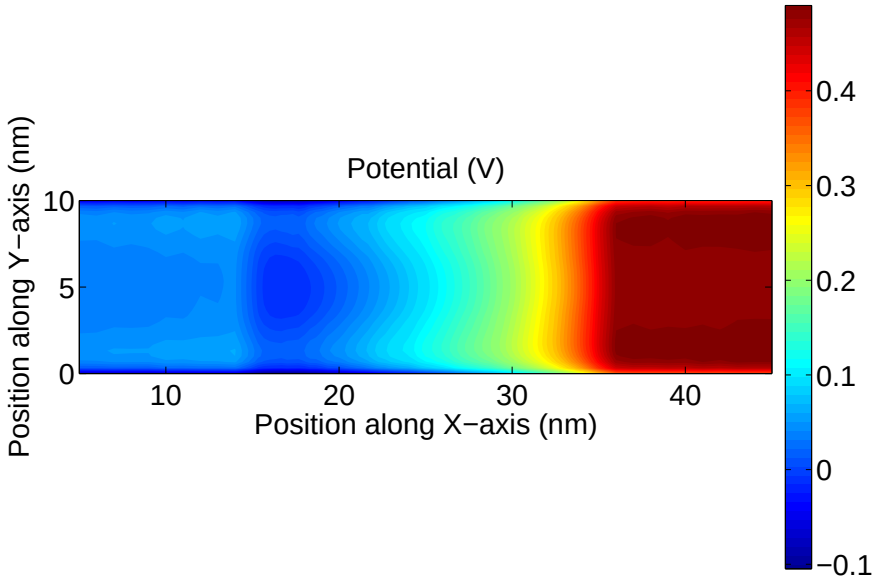


Figure 5.6: Potential contour from source to channel to drain (left to right).

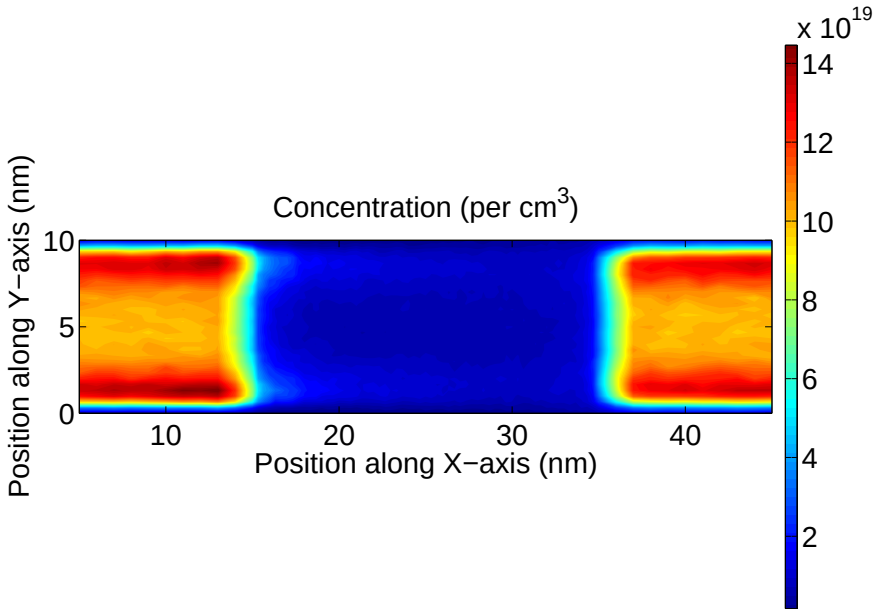


Figure 5.7: Charge concentration contour in the channel. Strong volume inversion starts at the source (left of the illustration) and continues throughout the channel length.

## 5.3 Double Gate versus Conventional MOSFET

The geometry of the conventional MOSFET used for comparison is shown in Figure 5.10. The dimensions are identical to that of DGFET except for one difference: the bottom gate and oxide are replaced with the substrate

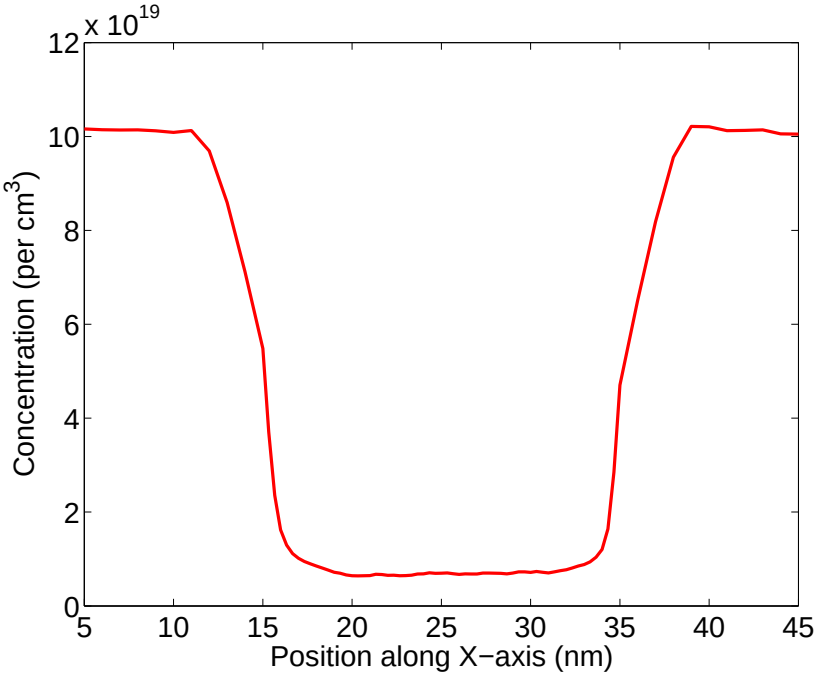


Figure 5.8: Charge concentration along the length of the device. Left edge represents source, followed by channel, and then drain.

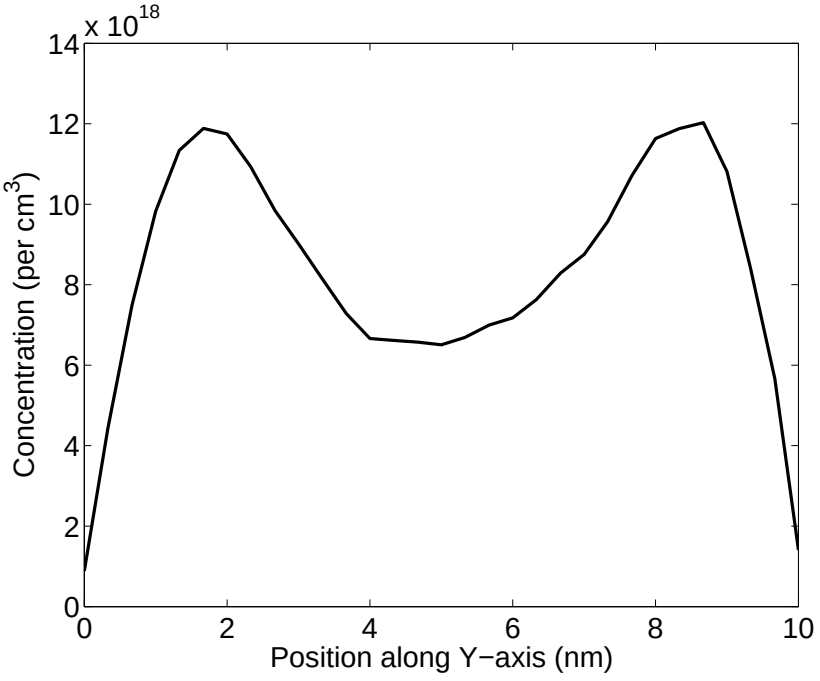


Figure 5.9: Charge concentration across channel thickness. Quantum correction modifies the shape of the concentration profile in short channels such as this.

by extending the channel material and a substrate contact is added at the bottom of the substrate. Both substrate and source are grounded so there is no body effect involved. Quantum correction is enabled for an accurate comparison with quantum-corrected DGFET simulations.

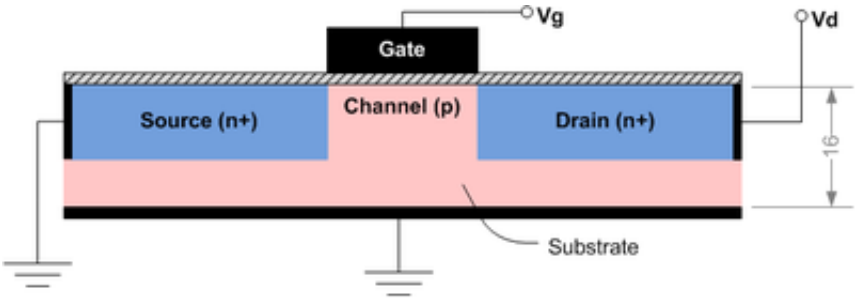


Figure 5.10: Conventional MOSFET geometry (to scale) used for comparison. Dimensions are the same as in the DGFET except that the bottom gate and oxide are replaced with substrate and substrate contact.

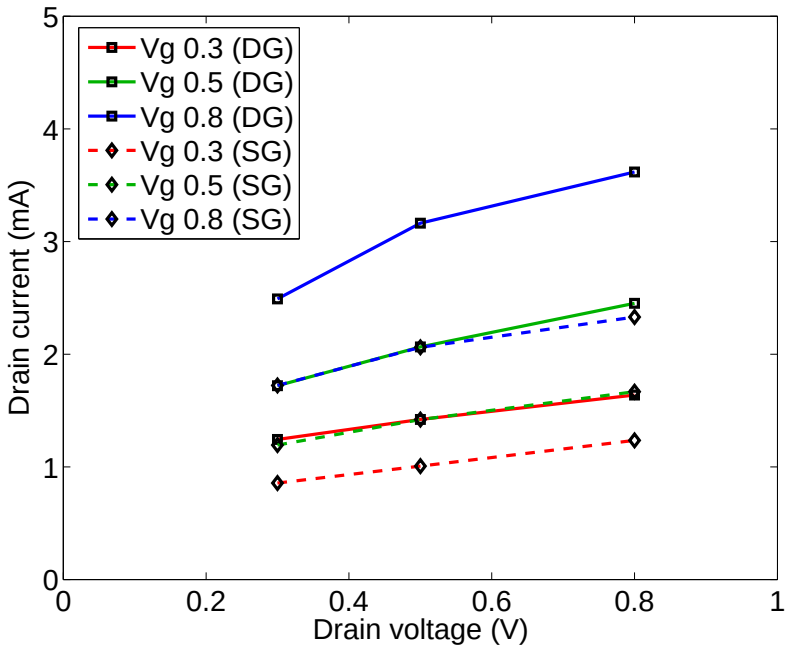


Figure 5.11: I-V characteristics of conventional and DGFET. Double gate MOSFET gives enhanced current drive because of better control of the channel by the gates.

The results of the simulations are shown in Figure 5.11. We can see that current drive in the case of DGFET is greater than that for conventional MOSFET resulting in better performance at nanometer scales.

## 5.4 High-$\kappa$ and $t_{ox}$

For simulation of different oxide dielectric constants and oxide thicknesses, the gate and drain bias was kept constant at 0.5 V. Oxide $\kappa$ was varied from 3.9 to 100 while oxide thickness was varied from 1 to 5 nm.

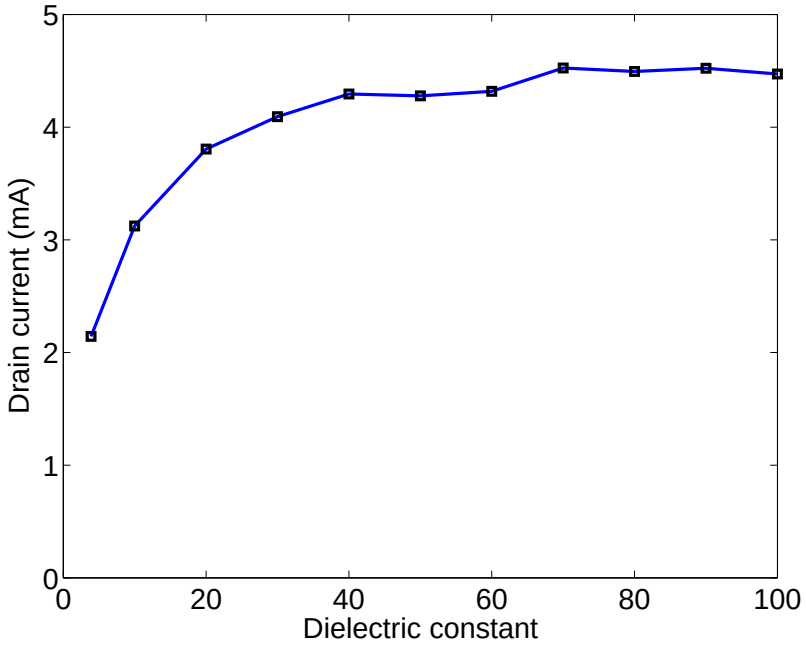


Figure 5.12: Drain current for different dielectric constants of the gate oxide. Gate and drain bias is fixed at 0.5 V.

The current variation for different $\kappa$'s is shown in Figure 5.12. As we can see, current does not increase much beyond a certain $\kappa$ around 40, so very high $\kappa$ dielectric materials might not give a significant advantage.

The current versus $t_{ox}$ variation is plotted in Figure 5.13. The current does not vary linearly with oxide thickness. On the other end, decreasing oxide thickness results in increased current drive. But this is only possible up to a point. It is believed that if the thickness is decreased beyond 1 nm, performance will suffer because of tunneling of carriers through such a small barrier, resulting in increased gate leakage.

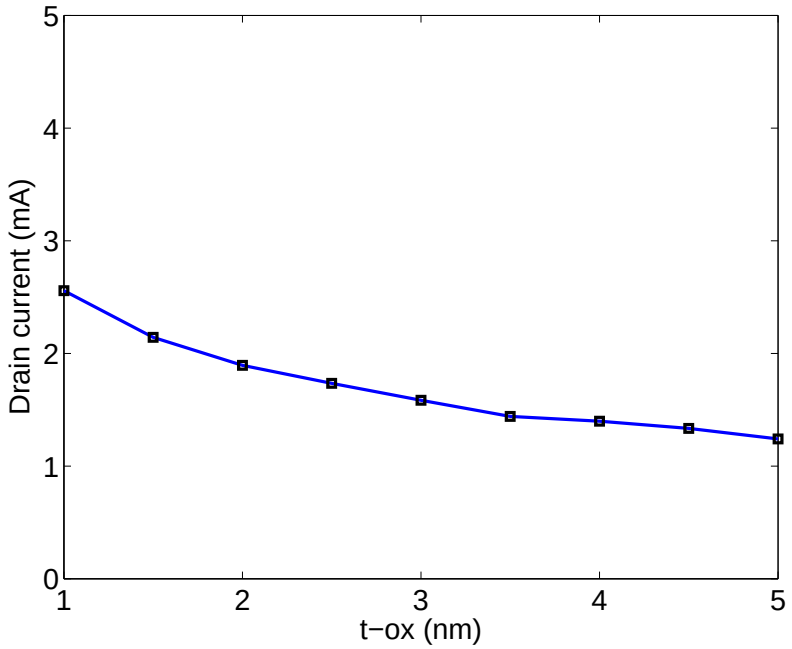


Figure 5.13: Drain current against different oxide thicknesses.

## 5.5 Misalignment Analysis

Performance of the device in the presence of misaligned gates with respect to each other and/or with respect to the channel was measured. For this purpose, a 9 by 9 matrix of gate positions was used. As an example, for row 5 in the matrix, the top gate was kept exactly on top of the channel, while the bottom gate was varied from close to the source edge all the way to close to the drain edge in increments of 4 nm. The example is illustrated in Figure 5.14. The contour of resulting currents for these 81 iterations is plotted as a contour in Figure 5.15. About half of these iterations are redundant because gates are symmetrical so it does not make a difference in current if the top gate is closer to the drain and the bottom gate closer to the source or vice versa. Such identical current values are visible in the plot in the form of symmetry across the off diagonal.

The middle of the plot corresponds to both gates exactly in the middle and perfectly aligned, i.e., zero misalignment for both gates. We have arrived at an interesting result here. The current is not maximum at zero misalignment. Instead, the maximum occurs when both gates are 4 nm into the source. In other words, there is an overlap of 4 nm with the source and 4 nm of underlap with the drain. Source and drain extension are known to

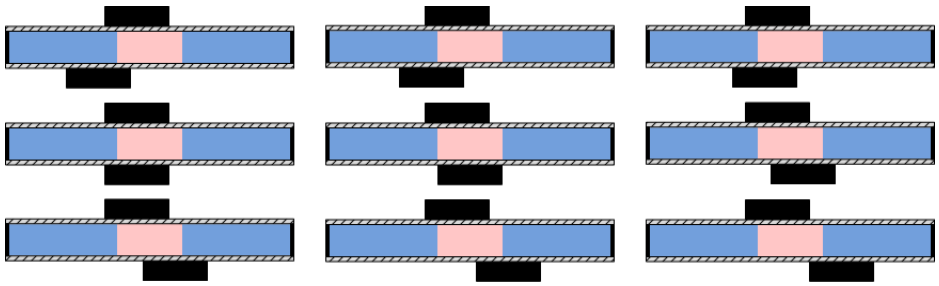

Figure 5.14: A set of 9 out of 81 simulations performed for misalignment analysis. Top gate is fixed in this set while bottom gate changes its location from -16 nm (from the channel center) to 16 nm in increments of 4 nm. The second configuration in the left column corresponds to the maximum in drain current value in the contour of Figure 5.15.

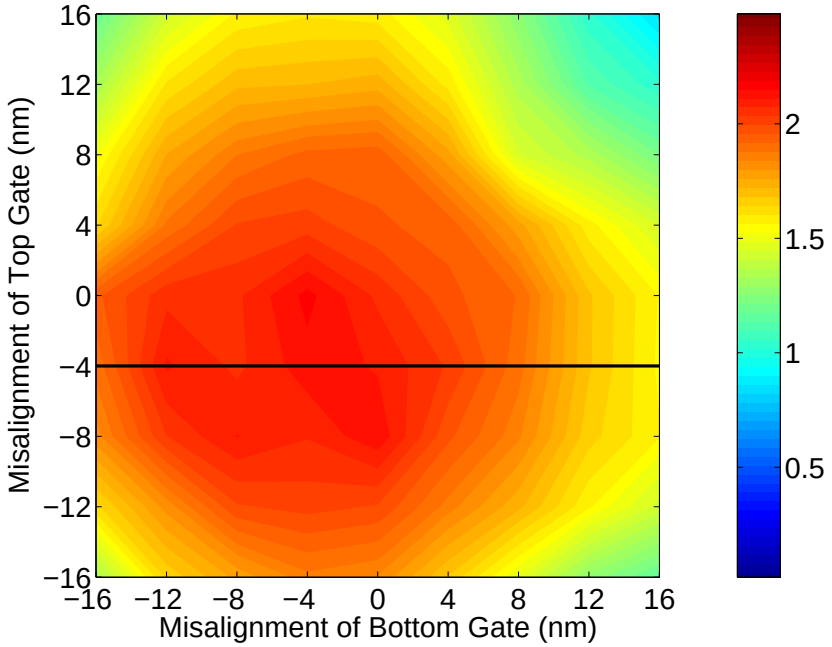


Figure 5.15: Drain current contour for all misalignment combinations. Black line at -4 nm below center corresponds to 9 values obtained with the configurations mentioned in Figure 5.14.

increase performance of the device up to a certain depth. But our result indicates that source extension is more dominant in increased current drive even if there is an underlap at the drain end. Figure 5.16 shows the drain current levels for four configurations of top gate misalignment.

In the worst cases, i.e, when most of the gate lengths are in drain with only a small overlap with the channel, current drops to only half of the maximum value. But since the current does not fall below 0.5 mA even for

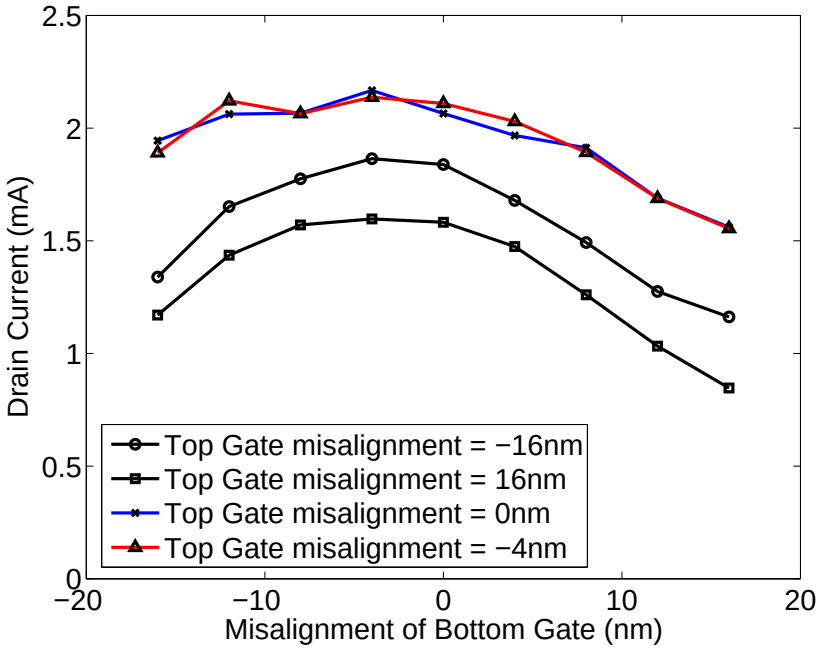


Figure 5.16: Current values for 4 sets of 9 misalignments. Lines with circle and square markers are the worst cases.

zero gate bias, we can subtract this much current from the results in order to fully appreciate the negative effects of misalignment on the current drive.

This work can be expanded in the future by analyzing other combinations of overlap and underlap, like increase and decrease in the length of one gate with respect to the other or with respect to the channel, and then varying the underlap and overlap of these gates with source and drain. As a result of that, a model can be developed to aid the design of these devices for optimum current drive. Overlap is not without disadvantages though, even if it gives higher current drive. Source and drain extensions have to be lightly doped in order to avoid lateral drain electric field and hence leakage current across the oxide [26]. Therefore further understanding of this phenomenon is needed to predict the optimum balance in performance.

# CHAPTER 6

# REVIEW OF THERMAL ANALYSIS OF ULTRASCALE DEVICES

This work analyzed the drain current performance as a result of variation of difference device parameters. Drain current is not the only metric for measuring performance of the device. With the ongoing scaling of devices into the nanometer regime, heating of the device has become a major concern. Performance of a device can be significantly affected if the heat dissipation is not properly managed. Heat management is a complex subject and extensive research on the topic is being carried out both in academia and in industry.

This chapter outlines a review of recent research in the area of thermal analysis at the ultrascale [27] and discusses possible future directions of DGFET that can result in better device design in relation to its thermal performance.

## 6.1 Thermal Effects in Ultrascale Devices

Thermal issues in high speed and small scale devices are one of the biggest concerns of the semiconductor industry as of today. The scaling trend that was following Moore's law for the last four decades has slowed down as a result. Power density of a chip has already approached that of a nuclear reactor, i.e., 100 W/cm$^2$ [28]. If this trend continues we shall soon see devices that will melt before they start operating.

Heat is primarily generated as a result of lattice vibrations, or phonons. In order to study thermal effects in ultrascale devices, phonon generation, annihilation, and transport need to be examined. This includes interaction of phonons with other particles, most notably electrons. In MoCa, electronic transport is studied through the electron BTE and the electronic band structure. In a similar fashion, phonon related phenomena can be

analyzed by considering the phonon BTE given by:

$$\frac{\partial N}{\partial t} + v(q)\frac{\partial N}{\partial x} = \delta N|_{el-ph} - \delta N|_{anharmonic} \tag{6.1}$$

where

$N$ : multi-dimensional distribution function (in x, q, and t)

$\delta N|_{el-ph}$ : change in distribution due to electron-phonon interaction

$\delta N|_{anharmonic}$ : change in distribution due to anharmonic decay

$q$ : k-space for phonon

$x$ : real space

$v(q)$ : phonon velocity

Anharmonic decay refers to the property of phonons that results in their interaction with each other when the crystal potential is not perfectly quadratic, i.e., the usual case [29]. As a result of this kind of interaction, two phonons collide and result in a third phonon whose momentum is the sum of that of the colliding phonons. We can extend this argument by considering an interaction of two phonons whose momenta are equal and opposite to each other. This results in annihilation of the phonons and it plays a crucial role in accounting for total heat distribution.

Similarly, consideration of phonon full band structure in the simulator can result in accurate calculation of phonon velocities which can be used in the phonon BTE to track particle trajectories and ultimately the paths of thermal current. Full phonon band structure for silicon is shown in Figure 6.1.

## 6.2 Possibilities of Improvement

MoCa provides a good opportunity for the exploration of electrothermal phenomena because of its accurate physics and reasonable simulation times. Phonon BTE support can be added to the simulator similar to the electronic BTE, and concentration of phonons can be tracked, which will

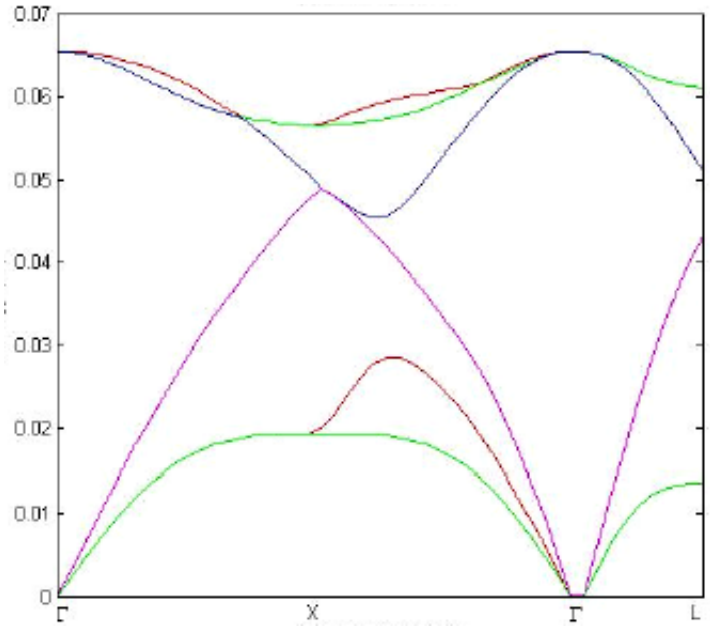


Figure 6.1: Phonon "full" band structure of Si.

give a better idea about the sites of heat generation, propagation, and dissipation.

DGFET geometry has oxide at top and bottom, and source and drain in front and back. This makes the channel a confined space. Silicon dioxide ($SiO_2$) is a poor thermal conductor and heat generated in the channel is trapped. Simulation that can examine the presence of heat in such confined geometry can be very beneficial to DGFET analysis. As a result of this analysis, the physical design space of the device can be explored and consequently the design of the DGFET can be optimized for both high dissipation and high drain current.

Other directions of this work include analysis using the 3-D version of MoCa. For example, we can analyze the performance of different DGFET geometries, such as FinFET versus planar. Misalignment analysis in 3-D MoCa will include overlap and underlap with the channel in the z-axis. Measurement of other performance parameters includes switching speed, leakage current, and power consumption. Different materials can be used in the source, drain, gate, and channel regions, and new heterostructures can be studied. New geometries can be explored. Results can be analyzed and calibrated against experimental values published by other groups. A threshold voltage model can be developed for the device. MoCa3D is a lot

more performance intensive than MoCa2D. As an example, a typical run of 0.2 million iterations of MoCa2D takes about one hour, whereas MoCa3D takes about 10 hours just to run 20,000 iterations, a difference of two orders of magnitude. There are some numerical algorithms that can potentially improve its performance. The conjugate-gradient algorithm can be replaced with a full multigrid method for the solution of nonlinear Poisson equation [30]. Parallelizing the code can provide speed improvements and make the model much more practical. Other improvements include calculation of hole concentration, successful convergence at or below the threshold voltage, simulation of time-dependent phenomena, and simulation of heterostructures. These speed improvements in the simulator will enable inclusion of new physics like consideration of phonons for heat analysis, and will help improve the accuracy of the results without increasing run time.

# CHAPTER 7

# CONCLUSION

Double gate MOSFET is a new class of devices that can help continue the scaling trend of the last 40 years. This work has reported increase in drain current of up to 60% compared to conventional MOSFET architecture of similar dimensions and the exact same materials. Short channel effects in the device are drastically reduced resulting in improved performance. Double gate MOSFET exhibits a new phenomenon of volume inversion that utilizes charge for transport in the volume of the devices instead of the edges, resulting in reduced surface scattering and increased mobility. The quantum effects and volume inversion manifest in a sub-50-nm channel. Drain current performance of a double gate MOSFET can be significantly affected as a result of variation of device parameters like oxide thickness, oxide dielectric constant, and misalignment of top and bottom gates with respect to each other as well as with the channel. Higher oxide dielectric constant increases drain current, but the current saturates beyond a value of 40. High-$\kappa$ dielectrics along with metal gates are already being used in today's devices. The importance of misalignment analysis is evident in benchmarking device performance. The current in the worst case scenario is 45% of the maximum. This indicates that, depending on the application, misalignment might be tolerable to a certain extent. Results show that source overlap due to misalignment increases the drain current even in the presence of drain underlap.

Computer simulation plays a very important role in semiconductor research and development. Monte Carlo offers the best balance of accuracy and performance at the mesoscopic scale, a scale that is relevant today and in the coming years. There are many possibilities for future work on this topic. Electrothermal analysis of the device, which will involve inclusion of effects due to the presence of phonons, will give insight into the heat performance of these structures. Knowledge of heat generation,

propagation, and dissipation can help improve the design of the device even further. The performance of the phonon-aware simulator can be enhanced by parallelizing the code, which will enable output of results in a reasonable time. Various geometries of the DGFET, along with different materials, can be explored to find the structures that can manage heat optimally.

# REFERENCES


[1] S. E. Thompson and S. Parthasarathy, "Moore's law: The future of Si microelectronics," *Materials Today*, vol. 9, no. 6, pp. 20–25, June 2006.

[2] M. Schulz, "The end of the road for silicon?" *Nature*, vol. 399, no. 6738, pp. 729–730, June 24 1999.

[3] "International technology roadmap for semiconductors 2008 update," 2008. [Online]. Available: http://www.itrs.net/Links/2008ITRS/Update/2008_Update.pdf

[4] G. E. Moore, "Cramming more components onto integrated circuits," *Proceedings of the IEEE*, vol. 86, no. 1, pp. 82–85, 1998.

[5] "International technology roadmap for semiconductors executive summary," 2001-2007. [Online]. Available: http://www.itrs.net/reports.html

[6] E. J. Nowak et al., "Turning silicon on its edge: Double gate CMOS/FinFET technology," *IEEE Circuits and Devices Magazine*, vol. 20, no. 1, pp. 20–31, 2004.

[7] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*, 3rd ed. Hoboken, NJ: Wiley-Interscience, 2007.

[8] J.-P. Colinge, *FinFETs and Other Multi-gate Transistors*. New York, NY: Springer, 2007.

[9] C. H. Lee, "Full band ensemble Monte Carlo simulation of silicon devices," Ph.D. dissertation, University of Illinois at Urbana-Champaign, Urbana, IL, 1994.

[10] G. A. Kathawala, "Monte Carlo simulation of nanostructures: Semiconductor devices to ion channels," Ph.D. dissertation, University of Illinois at Urbana-Champaign, Urbana, IL, 2005.

[11] K. Hess, *Advanced Theory of Semiconductor Devices*, 2nd ed. New York, NY: IEEE Press, 2000.

[12] R. M. Yorston, "Free-flight time generation in the Monte Carlo simulation of carrier transport in semiconductors," *Journal of Computational Physics*, vol. 64, no. 1, pp. 177–194, May 1986.

[13] M. Mohamed et al., "Size effects and performance assessment in nanoscale multigate MOSFET structures," *Journal of Computational and Theoretical Nanoscience*, vol. 6, no. 8, pp. 1927–1936, 2009.

[14] T. Sekigawa and Y. Hayashi, "Calculated threshold-voltage characteristics of an XMOS transistor having an additional bottom gate," *Solid-State Electronics*, vol. 27, no. 8-9, pp. 827–828, Sep. 1984.

[15] F. Balestra, S. Cristoloveanu, M. Benachir, J. Brini, and T. Elewa, "Double-gate silicon-on-insulator transistor with volume inversion: A new device with greatly enhanced performance," *IEEE Electron Device Letters*, vol. 8, no. 9, pp. 410–412, 1987.

[16] D. Hisamoto, T. Kaga, Y. Kawamoto, and E. Takeda, "A fully depleted lean-channel transistor (DELTA) - A novel vertical ultrathin SOI MOSFET," *IEEE Electron Device Letters*, vol. 11, no. 1, pp. 36–38, Jan. 1990.

[17] D. J. Frank, S. E. Laux, and M. V. Fischetti, "Monte Carlo simulation of a 30 nm dual-gate MOSFET: How short can Si go?" *Proceedings of IEEE International Electron Devices Meeting*, pp. 553–556, Dec. 13-16 1992.

[18] B. G. Streetman and S. Banerjee, *Solid State Electronic Devices*, 6th ed. Upper Saddle River, NJ: Pearson/Prentice Hall, 2006.

[19] M. Mohamed, "Monte Carlo simulation of non-local effects in electronic devices," M.S. thesis, University of Illinois at Urbana-Champaign, Urbana, IL, 2005.

[20] S. Deleonibus, *Electronic Device Architectures for the Nano-CMOS Era: From Ultimate CMOS Scaling to Beyond CMOS Devices*. Singapore: Pan Stanford; distributed by World Scientific, 2009.

[21] B. Winstead and U. Ravaioli, "A quantum correction based on Schrödinger equation applied to Monte Carlo device simulation," *IEEE Transactions on Electron Devices*, vol. 50, no. 2, pp. 440–446, 2003.

[22] B. A. Winstead, "Monte Carlo simulation of silicon devices including quantum correction and strain," Ph.D. dissertation, University of Illinois at Urbana-Champaign, Urbana, IL, 2001.

[23] F. Hassan, M. Chaabane, and C. Sathe, "Study of two-dimensional Schrödinger-Poisson solver," Oct. 2009, unpublished.

[24] A. Trellakis and A. T. Galick, "Iteration scheme for the solution of the two-dimensional Schrödinger-Poisson equations in quantum structures," *Journal of Applied Physics*, vol. 81, no. 12, p. 7880, June 15 1997.

[25] A. Trellakis, "Computational approaches to silicon-based nanostructures," Ph.D. dissertation, University of Illinois at Urbana-Champaign, Urbana, IL, 2000.

[26] J.-P. Colinge and C. A. Colinge, *Physics of Semiconductor Devices*. Boston, MA: Kluwer Academic Publishers, 2002.

[27] Z. Aksamija and U. Ravaioli, "Joule heating and phonon transport in silicon MOSFETs," *Journal of Computational Electronics*, vol. 5, no. 4, pp. 431–434, Dec. 1 2006.

[28] E. Pop, S. Sinha, and K. E. Goodson, "Heat generation and transport in nanometer-scale transistors," *Proceedings of the IEEE*, vol. 94, no. 8, pp. 1587–1601, 2006.

[29] C. Kittel, *Introduction to Solid State Physics*, 8th ed. Hoboken, NJ: Wiley, 2005.

[30] M. T. Heath, *Scientific Computing: An Introductory Survey*, 2nd ed. Boston, MA: McGraw-Hill, 2002.